

Paidéia, 2004, 14(29), 255-264

O USO DA AVALIAÇÃO E A AVALIAÇÃO DE SEU USO (ACERCA DA AVALIAÇÃO DA QUALIDADE DO ENSINO)¹

José Aparecido Da Silva²
FFCLRP – Universidade de São Paulo

Resumo: Nesse artigo são apresentados e discutidos os principais componentes de um programa de avaliação da qualidade ou excelência do ensino universitário conhecido como avaliação circular. Esses componentes são: 1) avaliação da qualidade do ensino através das opiniões e percepções dos estudantes, 2) avaliação pelos professores de suas relações e interações com os estudantes, e do ambiente acadêmico em que desenvolvem suas atividades de ensino e pesquisa, e 3) avaliação pelos professores das avaliações realizadas pelos seus estudantes (meta-avaliação). Os objetivos e as principais vantagens desse sistema de avaliação são apresentados e discutidos. O artigo conclui que, sob condições apropriadas, as escalas de avaliações dos estudantes sobre a qualidade do ensino são multidimensionais, confiáveis e estáveis.

Palavras-chave: Avaliação; Qualidade do ensino; Psicometria; Avaliação Acadêmica.

USE OF EVALUATION AND THE EVALUATION OF ITS USE (ON EVALUATION OF UNIVERSITY TEACHING QUALITY)

Abstract: This paper presents and discuss the main components of an evaluation program of the university teaching quality or teaching excellence, known as circular evaluation. These components are: 1) evaluation of teaching quality from student's judgments, 2) faculty evaluation of their relationship and interaction with students, and evaluation of the academic setting in which their teaching and research activities are carried out, and 3) faculty evaluation of their students' evaluations (meta-evaluation). The purposes and the main advantages of this evaluation system are presented and discussed. The article indicates that, under appropriate conditions, the scales for student's evaluation of teaching are multidimensional, reliable and stable.

Key-words: Evaluation; Evaluation of teaching quality; Psychometric; Academic Evaluation.

Examinando a evolução das ciências pode-se verificar que o problema da medida tem sido objeto de atenção dos cientistas desde a Antigüidade, a ponto de se considerar que uma ciência é mais evoluída quanto mais cedo foi sua preocupação com a mensuração. Por exemplo, a Física e, sobretudo, a Astronomia preocuparam-se desde a Antigüidade em medir a distância entre as estrelas e também em es-

tabelecer uma medida do tempo, seja em termos de dias, meses ou anos, ou mesmo de frações do dia como as horas e os minutos. Conhece-se também a importância fundamental das medidas nos trabalhos de Galileu (1564-1642) e de Kepler (1571-1630) e suas influências no desenvolvimento da mecânica racional durante o renascimento. Porém, foi Laplace (1749-1827) o primeiro a estudar sistematicamente os problemas teóricos levantados pela aplicação das medidas nas ciências de sua época. A Química permaneceu numa situação quase de estagnação até o momento em que Lavoisier (1745-1794) introduziu a balança analítica e através dela a mensuração. Indubitavelmente, foi este fato que permitiu uma rá-

¹ Artigo recebido para publicação em 12/07/2004; aceito em 13/10/2004

² Endereço para correspondência: José Aparecido Da Silva, Departamento de Psicologia e Educação, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, USP, Av. Bandeirantes, 3900, Monte Alegre, Ribeirão Preto, SP, 14040-901, E-mail: jadsilva@ffclrp.usp.br

pida evolução das ciências químicas. É claro que a ciência das medidas denominada de Metrologia nasceu e se desenvolveu no contexto das ciências mais velhas, as quais, graças a ela, naturalmente se chamam exatas. Atualmente, cada ciência tem seus problemas específicos a respeito das medidas e dos instrumentos de mensuração de que faz uso. Por exemplo, a Biologia tem o ramo da Biometria; na Economia encontra-se a Econometria que teve um desenvolvimento extraordinário nos últimos 20 anos; a Arqueologia se preocupou com a medida do tempo e através do uso das técnicas de datação por radiocarbono tornou possível determinar a idade de diferentes artefatos ou criações milenares dos homens; foi por meio da Arqueometria, que E. T. Hall, criou as ferramentas científicas usadas para desmascarar a fraude do Homem de *Pitldown* (suposto elo perdido entre o homem e o macaco) e determinar a idade do sudário de Turim. Na Sociologia e na Antropologia existem, respectivamente, a Sociometria e a Antropometria, que desempenham o mesmo papel que as disciplinas congêneres das outras ciências. Na Psicologia existe a Psicometria que trata das teorias, das técnicas, dos instrumentos, das escalas e dos diferentes modelos estatísticos e / ou matemáticos de avaliação e mensuração de atributos subjetivos.

A Psicometria, enquanto abordagem científica para a mensuração das características ou atributos de natureza psicológica, faz uso das escalas ou testes padronizados que são compostos de um conjunto de tarefas ou atributos administrados sob condições controladas e que são usadas para avaliar o conhecimento, as habilidades, os desempenhos e outras características psicológicas dos indivíduos; o seu campo teve início nos primeiros anos do século XIX, quando se tornou necessário diferenciar os indivíduos em função de suas características para objetivos clínicos, educacionais ou de trabalho. A avaliação psicológica é atualmente um dos campos mais amplos da Psicologia, da Educação e representa, talvez, uma das contribuições importantes da ciência psicológica/comportamental para a sociedade (Rust & Golombok, 1999). De fato, os testes, as escalas psicológicas e outras formas de avaliação, que não necessariamente a psicológica, são frequentemente utilizados nos processos envolvidos na avaliação humana. Praticamente todas as atividades são de alguma

forma avaliadas por um instrumento de medida (testes, escalas, inventários, questionários):- Somos avaliados nas escolas com o propósito de monitorar o desempenho; fazemos exames teóricos e práticos para a obtenção da carteira de motorista e para a habilitação ao exercício profissional, bem como com a finalidade de verificar alguma dificuldade de aprendizagem e também para sermos eventualmente premiados ou promovidos no trabalho em função de nossos desempenhos ou mesmo para serem corrigidas eventuais distorções em nossas funções (escalas de avaliação de desempenho ou da docência-ensino). A avaliação ocorre, pois, na escola, na universidade, na indústria, no trabalho (Da Silva, 2003).

A avaliação ou testagem psicológica envolve e afeta indivíduos, instituições e a sociedade como um todo. Fruto dessa importância, os psicólogos tentam fazer mensurações com o maior rigor possível, mas alguns testes ou escalas têm se tornado controversos pelo fato de as pessoas suspeitarem que são enviesados. Apesar da ampla variedade de aplicações e manifestações, as avaliações devem compartilhar um conjunto de propriedades comuns: serem fidedignas, válidas, padronizadas e livres de vieses. Há boas e más avaliações, e há uma ciência de como maximizar a sua qualidade, que é a Psicometria, campo da Psicologia com maior impacto na vida dos indivíduos.

Medida psicométrica

Um teste ou uma escala psicológica consiste numa situação experimental padronizada que serve de estímulo a dado comportamento ou construto que se pretende mensurar. Usualmente esse comportamento é avaliado mediante uma comparação estatística com os similares de outros indivíduos postos na mesma situação e, a partir daí, eles podem ser classificados quantitativa e qualitativamente segundo o construto mensurado. A escala ou o teste mede, mas, diferentemente do que ocorre com as medidas físicas, como comprimento e peso, há uma confusão considerável sobre o que, de fato, mede e quão bem faz isso.

O problema reside em que o que está sendo mensurado não é um objeto físico, e sim uma variável/construto interveniente, uma entidade hipotética

ou um traço latente. Por exemplo, ao se aferir uma escala para avaliação da qualidade e eficiência do ensino universitário, não se pode comparar diretamente o escore de um indivíduo (do professor avaliado pelo aluno) na escala, com o seu desempenho real ou verdadeiro; fica-se restrito a verificar como os escores das escalas/testes diferenciam indivíduos (professores) eficientes dos não eficientes em alguma dimensão subjacente ao processo de ensino-aprendizagem. Em outras palavras, está-se limitado a observar como os escores das escalas/testes diferenciam indivíduos (professores) eficientes dos não eficientes, de acordo com algumas dimensões (fatores que compõem a escala de avaliação) de como os indivíduos (professores) eficientes se comportariam. (Garcia & Sanchez, 1993; Sanchez, 1997; Greenwald, 1997; Marsh & Roche, 1997; D'Apollonia & Abrami, 1997; Greenwald & Gillmore, 1997; McKeachie, 1997).

A mensuração ou avaliação de construtos iguais à qualidade do ensino, eficiência do professor, ou mesmo da inteligência humana, é certamente limitada pela clareza com que se define o significado particular desses construtos, e isso tem se tornado um problema para os testes ou escalas que visam avaliar as dimensões ou fatores que compõem tais construtos. Assim, a rigor deve-se, em primeiro lugar, estabelecer, conhecer e definir (aferir a validade) as dimensões subjacentes ao construto que se pretende avaliar; sem isto, qualquer teste ou escala psicológica / educacional se torna inócua e não deve ser levada a sério.

Nesse sentido, o primeiro passo na construção de um instrumento (teste, escala, questionário, inventário) psicológico/educacional consiste em escrever uma definição detalhada do que ele busca mensurar. O processo deve ser iniciado com um grande número de itens carregados com os atributos que a escala pretende medir, testando-se, a seguir os itens com um grupo de pessoas (sujeitos: alunos, pacientes, clientes) que tenham características similares às das pessoas a que se destina; dessas aplicações preliminares, alguns itens serão retirados ou mantidos no conjunto da escala com base na sua capacidade de discriminação dos indivíduos (por exemplo, professores eficientes dos não eficientes). Posteriormente, técnicas estatísticas (análise fatorial, análise multidimensional, análise de regressão, correlação

canônica, correlação tetracórica, correlação bisserial, correlação ponto-bisserial, coeficiente Alfa de Cronbach, coeficientes de validade, coeficientes de fidedignidade) são aplicadas para agrupar ou estabelecer conglomerados de itens que se inserem em determinadas dimensões ou fatores que caracterizam o construto que se está avaliando e, também, para analisar as propriedades psicométricas da escala/teste. (Guilford, 1971; Pasquali, 2003).

Assim, uma escala de avaliação da qualidade do ensino por estudantes universitários poderia ser composta por várias dimensões (fatores) que hipoteticamente englobariam todo o processo de ensino-aprendizagem. Em adição, para propósitos comparativos, a unidade de análise teria as estimativas individuais dos estudantes, as médias das classes, os desvios padrão das médias, o que possibilitaria comparar o resultado de um indivíduo com os valores de referência (normas) e, conseqüentemente, até classificá-lo, o que permitirá conhecer sua posição, considerando o atributo ou construto que se está avaliando, comparativamente ao grupo tomado como padrão ou norma de referência, sendo a escolha da unidade de análise dependente do propósito da avaliação.

As escalas psicológicas, consideradas instrumentos de medida, devem possuir propriedades metrológicas que, em essência, são similares às requeridas em procedimentos de mensuração em outros domínios do saber. Portanto, elas são iguais aos instrumentos de medida de qualquer outra ciência, no sentido de que são feitas observações baseadas numa amostra pequena, mas cuidadosamente escolhida do comportamento do indivíduo. Sob esse aspecto, o avaliador procede como um químico que testa o teor carbônico de um composto a partir de uma pequena amostra, com a diferença de que, neste caso, a amostra consiste de uma série de comportamentos que se manifestam nas respostas dos indivíduos. (Carmines & Zeller, 1979; Spector, 1992).

O valor prognóstico ou diagnóstico de uma escala psicológica/educacional depende de seu grau como indicador do comportamento futuro do indivíduo ao se deparar com uma tarefa semelhante. Trata-se de generalizar a partir de resultados empíricos, como se faz em outras ciências. A propósito, assinala-se que os itens de uma escala não precisam assemelhar-se estritamente ao comportamento a ser pre-

visto por ela; é necessário, apenas, que haja uma correspondência empírica entre ambos (Da Silva, 2003; Pasquali, 2003).

As escalas psicológicas/educacionais têm propriedades. Dentro da Psicometria clássica, duas das mais importantes características de um teste são a fidedignidade (confiabilidade, precisão) e a validade: a) A fidedignidade refere-se à replicabilidade e à consistência dos escores da escala, contudo assinala-se que nenhuma escala será absolutamente consistente; não obstante, torna-se necessário que seja o mais fidedigna possível, contudo se ela não cumprir este requisito, torna-se impossível avaliar o seguinte, de tal modo que seria logicamente inviável que a escala fosse válida, mas não confiável. b) A validade diz respeito à relação entre os escores e os atributos psicológicos/educacionais reais que a escala está, supostamente, mensurando, sendo que este conceito refere-se genuinamente ao fato de o teste medir ou não os atributos do mundo real que pretende; nesse sentido, validade além de ser considerado o aspecto mais importante na avaliação de uma escala, é também a verdadeira medida de quão útil é o teste. O processo de validação de um teste é realizado acumulando evidências que suportam as inferências feitas de acordo com os resultados desta escala/teste, o que significa que não se está diante de um conceito unitário, havendo também várias formas de estimá-la. (Carmines & Zeller, 1979; Traub, 1994).

Avaliação institucional circular

A avaliação institucional circular refere-se a uma série inteira de avaliações realizadas pelos agentes da cadeia que fazem parte de um dado ambiente ou arranjo ensino-aprendizagem, nesse caso, a Universidade. Cada agente avalia os atributos (comportamentos, produtividade) que os outros agentes são capazes de desempenhar. Assim, os estudantes vêm a qualidade do ensino-instrução de seus professores, enquanto estes verificam a qualidade acadêmica global de seus estudantes (e não apenas os seus desempenhos). As instituições acadêmicas avaliam a qualidade do ensino, pesquisa e extensão de seus professores e, estes examinam o desempenho (eficiência) acadêmico da universidade e das agências de fomento à pesquisa (ou Ministério da Ciência e Tecnologia).

Um exemplo são as avaliações da CAPES ou mesmo o Provão e seus correlatos. Deste modo, cruzando-se as diferentes informações recolhidas a partir de várias fontes significativas em cada ambiente de ensino-aprendizagem, torna-se possível retratar o que ocorre na instituição (centro de pesquisa, departamento, instituto, faculdade, universidade).

De fato, na academia os construtos que são importantes para a tomada de decisão acerca da qualificação e da produtividade de seu corpo docente são o ensino, a pesquisa e a extensão (atividades administrativas também). Frequentemente, os métodos usados para se tomar decisões ao nível do corpo docente podem ser divididos em duas categorias: as atividades e realizações significativas dos docentes e as avaliações por pares dos docentes.

Grosso modo, as realizações dos docentes podem ser aferidas a partir do número de publicações, citações, qualidade delas (fatores de impacto), auxílios recebidos (número, montante recebido), apresentações de trabalhos em eventos (nacionais, internacionais), razão entre trabalhos apresentados em eventos e publicados. No ensino, as realizações seriam sumariadas pelas publicações ligadas a ele (livros didáticos, apostilas, demonstrações experimentais didáticas), formação de grupo de estudos-pesquisa e, também, avaliações pelos estudantes sobre a qualidade do ensino (indicador não usual no Brasil, pelo menos nas universidades públicas; mas em algumas privadas, utilizado de modo um pouco mais intenso, inclusive para remuneração e promoção, baseado em inferências). De outro lado, o conjunto de atividades de extensão, cultura e serviços parece não ter indicadores-critérios unânimes e nem todos concordam sobre a sua ponderação dentro da estrutura global da avaliação institucional. Certamente publicações ligadas à cultura, serviços, divulgações, atividades administrativas (carga horária), dados qualitativos sobre as contribuições à extensão e à cultura podem ser considerados.

O segundo componente usado para a avaliação do corpo docente tem sido a avaliação por pares. Certamente, no campo acadêmico esta acontece em 80 a 90 % do tempo, nos exames de candidatos em concursos públicos para ingresso na carreira, promoção na academia e efetivação. Como as outras medidas de desempenho acadêmico, as avaliações por

pares não têm um procedimento uniforme e são questionadas, pois os avaliadores dão apenas suas impressões enfocando: (1) realizações específicas num dado período, (2) relatório parcial descrevendo as atividades realizadas no período (curto ou longo) da carreira acadêmica ou (3) memorial completo, com as realizações relevantes da carreira acadêmica. Muitas dessas impressões são inferências acerca da qualidade do ensino, da eficiência do professor no processo interativo ensino-aprendizagem. (Sanchez & Mateo, 1992, 1993; Sanchez, Mateo, & Muñiz, 1996; Sanchez, 1997).

Escalas de avaliação da qualidade do ensino

Atualmente a prática de obter avaliações por parte dos estudantes acerca da qualidade do ensino universitário é firmemente estabelecida nos Estados Unidos e tem sido utilizada na maioria de suas universidades. Talvez, em decorrência desse sucesso, outros países desenvolvidos e em desenvolvimento têm começado a implementá-las em seu processo de avaliação institucional circular. (Garcia & Sanchez, 1993; Sanchez, 1997).

Há dois usos comuns para a informação obtida na aplicação das escalas de avaliação da qualidade do ensino realizadas pelos estudantes. Primeiro, a somativa que tem como objetivo possibilitar decisões administrativas sobre aumentos salariais, promoções, efetivação no cargo e mesmo incumbências de ensino. Segundo, a avaliação formativa em que os dados coletados são usados com objetivo de melhorar as atividades de ensinar, ou seja da didática e da eficiência do professor, o que pode, certamente, afetar decisões sobre promoção ou efetivação, mas sua intenção primária é o desenvolvimento do docente. (Sanchez, 1997; Marsh & Roche, 1997; McKeachie, 1997).

Esses dois tipos de avaliação requerem evidências diversas. A somativa procura identificar diferenças na qualidade entre docentes e que devem ser baseadas em dados bem confiáveis. Para esse propósito o uso das avaliações dos estudantes, fundadas nas respostas médias das classes para as respostas globais, são as mais apropriadas. De outro lado, as questões globais são pouco úteis na avaliação formativa, e aí o “como” da eficácia do ensino é o

aspecto central. O “feedback” do estudante sobre comportamentos específicos do ensino é um componente necessário. – outras fontes de “feedback”, como video tapes, observações, comentários feitos pelos pares, consultas e mesmo auto-avaliações, são também estudadas. Reafirmando: “*Complete agreement on everything is not necessary to make a concept meaningful*”. (Greenwald, 1997).

Algumas das razões alegadas para avaliar o ensino são:

- 1) Comprovar sua excelência para a promoção ou efetivação e, neste caso as avaliações podem servir como indicadores do valor e validade do que está sendo feito e ensinado; indiretamente espelham a qualidade do ensino, sendo que os próprios docentes são requeridos a submeter uma descrição de suas atividades de ensino.
- 2) Diagnosticar potencialidades e fraquezas. Aqui as avaliações dos estudantes são vistas como componentes importantes no desenvolvimento acadêmico, possibilitando enriquecer a qualidade do ensino, a aprendizagem e elas podem ser obtidas durante a disciplina, próximas ao fim do semestre (ano), ao término de um módulo (unidade ou bloco) de conteúdo, para o propósito de confirmar as potencialidades do ensino e identificar as fraquezas de modo a fornecer melhor ambiente de aprendizagem aos estudantes no futuro.
- 3) Melhorar a qualidade da disciplina e do ensino. Descobrir como os estudantes experienciam seu ambiente de aprendizagem é certamente um primeiro passo para cumprir este objetivo. O delineamento do conteúdo, as estratégias de avaliação, a disponibilidade de recursos didático/pedagógicos, a integração entre conteúdos são elementos importantes e válidos a serem considerados num programa completo de avaliação, de refinamento e de melhoramento da qualidade do ensino.
- 4) Engajar os estudantes mais ativamente no processo de ensino-aprendizagem. As avaliações dos estudantes fornecem um meio de comunicação entre eles e o professor, o que pode levar a um envolvimento que eleve o nível de instrução, estimulando a instituição a considerar seus objetivos e metas gerais e mesmo os seus valores.
- 5) Fornecer evidências para um exame da qualidade institucional. Nesse caso, avaliações sistemáticas

pelos estudantes dão informações sobre instrutores e disciplinas com o propósito de que esses selecionem disciplinas e professores, ou haja o encorajamento para o melhoramento institucional, aumentando as chances de que a excelência no ensino seja reconhecida e valorizada (recompensada também).

- 6) Investigar as dificuldades dos estudantes. Os resultados obtidos pelas escalas de avaliação da qualidade do ensino podem indicar quais as reais dificuldades de aprendizagem dos estudantes, em quais contextos, situações e de que modo aparecem, permitindo a correção de rotas para facilitar o processo de aprendizagem.
- 7) Satisfação profissional. Resultados de avaliações da qualidade do ensino pelos alunos são base para decisões administrativas e, aumentam a probabilidade de que a qualidade do ensino seja reconhecida, valorizada, e que bons professores sejam efetivados. O reforço social de avaliações favoráveis dá incentivos adicionais para a melhora do ensino, mesmo com professores já efetivos. Há quem registre que o “feedback” dos alunos é útil para dirigir esforços visando o aperfeiçoamento de suas técnicas de instrução.
- 8) As avaliações da qualidade do ensino pelos estudantes podem ser úteis para monitorar as suas inovações, especialmente no uso de tecnologias educacionais que visam prover ganho no processo ensino-aprendizagem.

Esse movimento em relação à avaliação do ensino pelos estudantes, com sua curta história, mas enorme vitalidade, força e interesse, tem fornecido um número considerável de instrumentos (escalas, questionários, inventários). A maioria deles tenta medidas das variáveis ditas relevantes para um bom ensino: a) estruturação e a organização do material pelo professor; b) coerência e clareza das explicações; c) habilidades motivacionais; d) interação positiva entre professor e estudantes. O número de dimensões (fatores) resultantes da análise fatorial dos itens desses instrumentos usualmente varia entre dois e nove; tais resultados parecem indicar uma característica multidimensional do construto qualidade do ensino. (Garcia & Sanchez, 1993; Marsh & Roche, 1997).

Uma das escalas freqüentemente utilizada em universidades americanas, nas da Espanha, Canadá,

Austrália, Nova Zelândia, Hong Kong e Papua Nova Guiné, tem revelado que nove fatores ou dimensões ligadas à eficiência do ensino podem agrupar praticamente todos os itens usados para mensurar a qualidade do processo ensino-aprendizagem. Elas são: 1) Aprendizagem/valor (Eu achei essa disciplina intelectualmente estimulante e desafiadora); 2) Entusiasmo (O estilo de apresentação do professor manteve meu interesse durante a aula); 3) Organização (Os materiais utilizados pelo professor foram bem preparados e cuidadosamente explicados); 4) Interação de Grupo (Os estudantes foram encorajados a participar das discussões de classe, a expressar suas próprias idéias); 5) Interação Individual (O professor fazia os estudantes se sentirem bem-vindos ao buscar ajuda/orientação dentro ou fora da sala de aula); 6) Amplitude (por ex., O professor contrastava as implicações de várias teorias e adequadamente discutia os recentes desenvolvimentos na área); 7) Exames/Provas (Os métodos de avaliação das tarefas dos estudantes testavam o conteúdo da disciplina tal como enfatizado pelo professor); 8) Tarefas/Exercícios (O material de leitura/textos foi valioso, contribuindo para uma apreciação e compreensão do assunto); 9) Carga de Trabalho/Dificuldade (fácil - difícil; leve-pesada; lento-rápido) (Kremer, 1990; Abrami, D'Apollonia, & Cohen, 1990; Cranton, & Smith, 1990; Marsh, Hau, Chung, & Siu, 1997).

Ao lado dessas nove dimensões a escala inclui uma avaliação global da qualidade do ensino considerando juntos todos os itens de cada dimensão (por ex., avaliação geral da disciplina e avaliação geral do professor; comparada com outras disciplinas que tenho cursado posso dizer que essa disciplina é: muito fraca, fraca, na média, boa, muito boa; comparado com outros professores de outras disciplinas, o professor dessa disciplina é: péssimo, ruim, mediano, bom, ou excelente). Em adição, variáveis como o tamanho da classe, a série (semestre-ano), idade e sexo do professor, são anotadas. Os estudantes são instruídos a responder aos itens numa escala de 9 pontos (tipo Likert) cujos escalões (categorias de resposta) variam de 1 (discorda totalmente) a 9 (concorda totalmente). O número de pontos em escalas dessa natureza pode também variar de 1 a 5 ou entre 1 e 7, sem perder informação relevante. Questões abertas permitindo comentários ou sugestões podem ser co-

locadas ao final das escalas. (Antonak & Livneh, 1988; Engeldrum, 2000).

Torna-se importante mencionar que as características psicométricas dessas escalas têm sido regular e profundamente analisadas, indicando altos coeficientes de validade, de fidedignidade, poder de discriminação entre professores bons (eficientes) e péssimos (ineficientes). Do mesmo modo, variáveis tais como o tamanho da classe, o sexo e idade do professor e sua interação com o sexo e idade dos estudantes, tempo do dia em que a disciplina é ministrada, natureza da disciplina (obrigatória, optativa) e tipo de curso (exatas, humanas, biológicas), nível acadêmico dos estudantes, grau de indulgência na atribuição das notas (*Dr. Fox Effect*), nível da disciplina (graduação ou pós-graduação - há escalas específicas para a avaliação da supervisão de pesquisa - mestrado ou doutorado - por parte dos estudantes de pós-graduação envolvendo as dimensões concordância e satisfação com a supervisão) e outros efeitos específicos de uma dada disciplina (cursos) têm sido sistematicamente investigadas e quanto da variância cada uma dessas variáveis explica na avaliação global do professor e da disciplina. Uma coisa é certa, a despeito das inevitáveis flutuações nesse campo de pesquisa: para um dado professor, ensinando grosso modo o mesmo tipo de disciplina de ano para ano, as avaliações feitas pelos estudantes são medidas extremamente confiáveis e válidas da qualidade do ensino desse professor. E, uma adequada combinação das estimativas dos estudantes com outras formas de “feedback” e avaliações, certamente permitirá os ajustamentos necessários tanto das estratégias quanto das técnicas de ensino. Finalmente, esta mesma escala contendo o conjunto de dimensões (fatores) acima descrito pode também ser utilizada pelo professor como uma escala de auto-avaliação. Para isso, nos itens acima exemplificados deve-se mudar apenas o sujeito da ação ou do responsável pelo comportamento implicado em cada um deles. Como disse, certa vez, Winston Churchill, mas num diferente contexto: “*Give us the tools and we’ll finish the job!*”

Com esses dados em mãos, cada unidade educacional pode iniciar um período de reflexão com o objetivo de descobrir as razões para as eventuais discrepâncias entre os resultados obtidos a partir desses dois tipos de avaliações formativas (dos estudantes e

auto-avaliação por parte do professor). Uma análise desses dados pode servir para a tomada de decisões apropriadas visando melhorar a qualidade do ensino, algo que é sempre possível.

Avaliação do ambiente acadêmico por parte dos professores

Mesmo que a avaliação da qualidade do ensino tenha sido feita pelos estudantes quanto pelos professores de forma válida e fidedigna, a imagem do processo ensino-aprendizagem dentro da universidade continua a ser parcial e, assim, bastante pobre. Por esta razão torna-se necessário introduzir novos dados avaliativos considerando o ambiente em que esse processo se realiza. A principal fonte de informação será o corpo docente, com objetivos altamente diversos sob avaliação.

O primeiro objetivo focaliza ainda os estudantes, mas não mais em termos de seu desempenho e sim em termos de suas relações acadêmicas com o professor. Alguns dos atributos (itens) que podem ser avaliados são: a participação dos estudantes na classe, o interesse mostrado em relação ao assunto, o tipo de trabalho acadêmico que eles realizam. Considerando essa dimensão, a natureza circular do modelo de avaliação institucional torna-se perfeitamente clara. Isto se alia às vantagens de usar as informações cruzadas derivadas de duas fontes de informação altamente importantes em qualquer sistema educacional: a dos professores e a dos estudantes.

O segundo objetivo é avaliar o clima intradepartamental. Nesse sentido, como os estudantes avaliam seus professores dentro de um contexto particular, também se torna desejável e conveniente que o professor tenha oportunidade de realizar sua própria avaliação do ambiente em que ele desempenha suas atividades e/ou funções de ensino, pesquisa e extensão (serviços). Assim, a avaliação do ambiente acadêmico por parte dos professores poderá servir de linha de base sobre a qual programas de intervenção claramente delineados podem ser implementados visando à promoção ou o melhoramento da produção científica e desenvolvimento profissional dos docentes. Para isso, as escalas utilizadas têm-se limitado a focalizar diferentes facetas (dimensões) de um amplo campo que é o que mais in-

fluencia a produtividade de pesquisa e do ensino do corpo docente: (a) clima organizacional, (b) satisfação do docente no exercício de suas funções, (c) mecanismos e procedimentos de seleção e promoção, (d) formas de remuneração - salários, (e) o ambiente psicossocial, e (f) desempenho dos serviços de apoio institucional. Estes aspectos contextuais constituem o que denominamos: vitalidade do corpo docente e qualidade da vida universitária. Exemplos de itens empregados para mensurar estas dimensões são: As condições materiais nas quais eu desempenho o meu trabalho são satisfatórias; Economicamente tem sido possível desenvolver os meus projetos de pesquisa; A instituição universitária estimula-me a melhorar as minhas atividades de docência; Há critérios claros para avaliar as atividades de pesquisas; Eu recebo apoio de meus colegas nas atividades que realizo; Há um nível satisfatório de comunicação acadêmica entre os membros de meu departamento; Minhas relações com os meus colegas de departamento favorecem minhas atividades acadêmicas; Há uma boa cooperação intradepartamental no desenvolvimento dos projetos de pesquisas; Desempenho da biblioteca; Desempenho dos serviços de secretaria; Desempenho da Reitoria e Pró-reitorias; Desempenho dos sistemas de áudio-visual-computação. A tarefa dos professores consiste em avaliar cada um dos itens que compõem essa escala usando escores que variam de 1 a 7 (modelo tipo Likert) indicando o grau de concordância com o conteúdo de cada um deles. As características psicométricas analisadas nessas escalas revelam que elas são fidedignas, válidas e com altos índices de homogeneidade e validade dos itens que a compõem. Em outras palavras, é um bom instrumento para avaliar a qualidade do ambiente acadêmico. Com estes dados em mãos, certamente a Reitoria, as Pró-reitorias da Universidade e até mesmo a Diretoria de uma Unidade podem avaliar melhor suas responsabilidades e mesmo implementar políticas acadêmico-científicas objetivando a melhoria da qualidade do ensino, da pesquisa, dos serviços de extensão e, principalmente, dos órgãos de serviços de apoio à vida universitária. Em outras palavras podem melhorar a qualidade de vida *universitária*. (Garcia & Sanchez, 1993; Sanchez, 1997).

Outros componentes da avaliação institucional circular

Ao lado dos componentes pertinentes à avaliação da qualidade do ensino (por estudantes e por professores) e à avaliação do ambiente acadêmico (por parte dos professores), há nessa cadeia circular de avaliação do processo ensino-aprendizagem dois outros componentes: (1) a avaliação dos professores pelos administradores da universidade; nesse caso pode ser conduzida pelo Reitor, Pró-reitores e Diretores de uma dada instituição. Ela se caracteriza por ser voluntária, ocorre após períodos fixos (por ex. cinco anos de contrato; período probatório no RDIDP) e é fundamentalmente somativa (aditiva), isto é, pode implicar a renovação ou não do contrato de trabalho ou mesmo alterações na carreira acadêmica - mudança de regime, efetivação (alteração nas funções ou cargos). Importante mencionar que em algumas universidades americanas e européias está se começando a introduzir as avaliações dos estudantes da qualidade do ensino como um dos critérios a considerar quando da renovação do contrato ou alterações de função. Enquadram-se também aqui as avaliações periodicamente realizadas pelas agências de fomentos, tais como a CAPES quando avaliando os programas de pós-graduação de forma somativa e formativa, e pelo MEC objetivando verificar as condições de ensino-pesquisa das instituições universitárias. (2) Finalmente, no último componente denominado de meta-avaliação deve-se destacar que aos professores é dada a possibilidade de avaliar a utilidade das avaliações feitas pelos estudantes acerca de sua qualidade de ensino-instrução, tanto com respeito ao aspecto formativo (melhoramento na qualidade do ensino) quanto ao aspecto somativo (considerando uma futura promoção, mudança de regime, aumento de salários).

Uma vez que estes quatro componentes da cadeia avaliativa estejam completos e adequadamente analisados, tanto do ponto de vista estatístico quanto comparativo, a avaliação institucional circular pode ser finalizada. Este é um ciclo em que cada agente da cadeia avaliativa do complexo processo de ensino-aprendizagem avalia os outros e ao mesmo tempo está sendo avaliado pelas suas contrapartes. O sistema deve ser transparente, e usar critérios e escalas

bem definidas, válidas, fidedignas e que englobam as diferentes dimensões (facetas) do processo de ensino-aprendizagem, ou seja, sejam capazes de objetivamente avaliar a dinâmica e as características multidimensionais da vida acadêmica - universitária. Importante concluir que o avaliador (professor ou estudante) é meramente o mensageiro de um dinâmico e complexo processo. Assim: *Don't shoot the Messenger, he is doing his best!* (Greenwald, 1997).

Centro de estudos e pesquisas em ensino-aprendizagem (CEPEA)

Considerando que: (1) o processo de avaliação da qualidade do ensino envolve a construção de escalas (questionários, inventários) cujas propriedades metrológicas necessitam ser aferidas com bastante rigor do ponto de vista da teoria geral da mensuração, e (2) entendendo que muitas das universidades públicas brasileiras estão apenas se iniciando e têm pouca experiência nessa dinâmica quando comparadas com as congêneres americanas e européias que possuem longa tradição nesse processo de avaliação, eu recomendaria que as pró-reitorias de graduação e a de pós-graduação constituíssem um centro especializado na construção, aplicação e análise das características psicométricas de escalas que objetivassem avaliar – mensurar as múltiplas dimensões subjacentes ao processo de ensino-aprendizagem. Acredito que um trabalho a ser realizado em longo prazo, psicometricamente bem desenvolvido, transparente, dinâmico e contando com a colaboração dos diferentes segmentos que compõem a comunidade acadêmica, poderia ser útil, e promissor para o desenvolvimento e melhoramento da qualidade de ensino das universidades públicas brasileiras. Com disse Claude Bernard: *"In ignorance, abstain."*

Referências Bibliográficas

- Abrami, P. C., D'Apollonia, S. & Cohen, P.A. (1990). Validity of student ratings of instruction; What we know and what we do not. *Journal of Educational Psychology*, 82(2), 219-231.
- Antonak, R.F. & Livneh, H. (1988). *The measurement of Attitudes toward People with Disabilities: Methods, Psychometrics and Scales*. Springfield: Charles C. Thomas.
- Carmines, E.G. & Zeller, R.A. (1979). *Reliability and Validity Assessment*. Thousand Oaks: Sage.
- Cranston, P. & Smith, R. A. (1990). Reconsidering the unit of analysis: A model of student ratings of instruction. *Journal of Educational Psychology*, 82(2), 207-212.
- D'Apollonia, S. & Abrami, P.C. (1997). Navigation student ratings of instruction. *American Psychologist*, 52(11), 1198-1208.
- Da Silva, J.A. (2003). *Inteligência Humana: Abordagens Biológicas e Cognitivas*. São Paulo: Lovise.
- Engeldrum, P.G. (2000). *Psychometric Scaling*. Winchester: Imcotek Press.
- Garcia, M.A.M. & Sanchez, J.F. (1993). Dimensões de la calidad de la enseñanza universitária. *Psicothema*, 5(2), 265-275.
- Greenwald, A.G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52(11), 1182-1186.
- Greenwald, A. G. & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist*, 52(11), 1209-1217.
- Guilford, J. P. (1971). *Psychometric Methods*. New York: McGraw-Hill.
- Kremer, J. F. (1990). Construct validity of multiple measures in teaching research, service and reliability of peer ratings. *Journal of Educational Psychology*, 82(2), 213-218.
- Marsh, H. W., Hau, K. T., Chung, C. M. & Siu, T. L. P. (1997). Students' evaluations of university teaching: Chinese version of the students' evaluations of educational quality instrument. *Journal of Educational Psychology*, 89(3), 568-572.
- Marsh, H. W. & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and

- utility. *American Psychologist*, 52(11), 1187-1197.
- McKeachie, W. J. (1997). Student ratings: The validity of use. *American Psychologist*, 52(11), 1218-1225.
- Pasquali, L. (2003). *Psicometria: Teoria dos testes na psicologia e educação*. Rio de Janeiro: Editora Vozes.
- Rust, J. & Golombock, S. (1999). *Modern Psychometric: The Science of Psychological Assessment*. New York: Routledge.
- Sanchez, J. F. (1997). A Spanish evaluation model in higher education: Circular evaluation. *Higher Educational Management*, 9(1), 71-84.
- Sanchez, J. F. & Mateo, M. A. (1992). Student evaluation of university teaching quality: Analysis of a questionnaire for a sample of university students in Spain. *Educational Psychological Measurement*, 52, 675-686.
- Sanchez, J. F. & Mateo, M. A. (1993). The development and factorial validation of the academic setting evaluation questionnaire. *Educational Psychological Measurement*, 53, 425-435.
- Sanchez, J. F., Mateo, M. A. & Muñiz, J. (1996). Valoración por parte del profesorado de la evaluación docente realizada por los alumnos. *Psicothema*, 8(1), 167-172.
- Spector, P. E. (1992). *Summated Rating Scale Construction*. Thousand Oaks: Sage.
- Traub, R. E. (1994). *Reliability for the Social Sciences: Theory and Applications*. Thousand Oaks: Sage.

NT:**Conferência apresentada no 1º Seminário Internacional de Avaliação da Graduação e da Pós-graduação, organizado pelas Pró-reitorias de Graduação e de Pós-graduação da Universidade de São Paulo, em 28 de novembro de 2003, São Paulo, Capital. Agradecimentos ao Prof Dr. Juan Fernández Sanchez, da Universidad Complutense de Madrid, Espanha, pelas sugestões e discussões acerca da natureza multidimensional das escalas de avaliação da excelência do ensino.